

TECHNICKÁ UNIVERZITA V LIBERCI
Fakulta mechatroniky, informatiky a mezioborových studií

Studijní program: N2612 – Elektrotechnika a informatika
Studijní obor: 1802T007 - Informační technologie

**Možnosti hlasové interaktivní komunikace pomocí
bezdrátových technologií v robotických aplikacích**

**The options of interactive voice communications via
wireless technologies in robotic applications**

Diplomová práce

Autor: **Issatay Iskakov**
Vedoucí diplomové práce: Ing. Miroslav Holada, Ph.Dr.

V Liberci 3. 1. 2013

ZADÁNÍ DIPLOMOVÉ PRÁCE

(PROJEKTU, UMĚLECKÉHO DÍLA, UMĚLECKÉHO VÝKONU)

Jméno a příjmení: Bc. Issatay Iskakov
Osobní číslo: M10000227
Studijní program: N2612 Elektrotechnika a informatika
Studijní obor: Informační technologie
Název tématu: Možnosti hlasové interaktivní komunikace pomocí bezdrátových technologií v robotických aplikacích
Zadávací katedra: Ústav informačních technologií a elektroniky

Z á s a d y p r o v y p r a c o v á n í :

1. Seznamte se se současným stavem hlasového ovládání v robotických aplikacích především na pracovišti školitele. Dále prostudujte přenosové možnosti stávajících bezdrátových technologií (např. radiové spojení, Wi-fi, BT ...).
2. Navrhněte jednoduchou interaktivní hlasovou úlohu pro potřeby testování vybraných bezdrátových technologií.
3. Jednotlivé technologie do úlohy implementujte.
4. Proveďte testování a na základě dosažených výsledků formulujte stávající možnosti pro dané využití.

Rozsah grafických prací:

Dle potřeby dokumentace

Rozsah pracovní zprávy:

cca 40-50 stran

Forma zpracování diplomové práce: tištěná/elektronická

Seznam odborné literatury:

- [1] HLAVÁČ V., SEDLÁČEK M.: Zpracování signálu a obrazu, Skripta FEL
ČVUT, Praha 2000, ISBN 80-01-02114-9

Vedoucí diplomové práce:

Ing. Miroslav Holada, Ph.D.


Ústav informačních technologií a elektroniky

Datum zadání diplomové práce:

1. října 2012

Termín odevzdání diplomové práce:

20. května 2013


prof. Ing. Václav Kopecký, CSc.
děkan




prof. Ing. Zdeněk Plíva, Ph.D.
pověřen vedením ústavu

V Liberci dne 1. října 2012



Prohlášení

Byl jsem seznámen s tím, že na mou diplomovou práci se plně vztahuje zákon č. 121/2000 o právu autorském, zejména § 60 (školní dílo).

Beru na vědomí, že TUL má právo na uzavření licenční smlouvy o užití mé diplomové práce a prohlašuji, že **s o u h l a s í m** s případným užitím mé diplomové práce (prodej, zapůjčení apod.).

Jsem si vědom toho, že užít své diplomové práce či poskytnout licenci k jejímu využití mohu jen se souhlasem TUL, která má právo ode mne požadovat přiměřený příspěvek na úhradu nákladů, vynaložených univerzitou na vytvoření díla (až do jejich skutečné výše).

Diplomovou práci jsem vypracoval samostatně s použitím uvedené literatury a na základě konzultací s vedoucím diplomové práce.

Datum: 3. leden 2013

Podpis



PODĚKOVÁNÍ

Rád bych tímto poděkoval panu Ing. Miroslavu Holadovi za odborné vedení a za konzultace při zpracovávání diplomové práce.

Dále bych chtěl poděkovat všem, kteří mě podporovali během realizace diplomové práce.

Issatay Iskakov



ANOTACE

Cílem této práce bylo prostudovat možnosti hlasové komunikace mezi člověkem a robotem, stanovit nejlepší způsob přenosu audio dat po bezdrátových sítích a zjistit jejich prodlevu oproti přímé komunikaci bez bezdrátového přenosu. Bylo použito k prozkoumání tři typů bezdrátových technologií – radiové spojení, Wifi a Bluetooth. Práce byla zaměřena na změření zpoždění, t.j. kolik času uplyne při přenosu audio dat po bezdrátových sítích s jejich využitím především v robotických aplikacích. Primárně bylo potřeba vytvořit aplikaci, která rozpoznává lidskou řeč a pokud najde náležitou odpověď na hlasový povel v zásobě slov, vykoná další krok hlasového dialogu. Toto vedlo k vytvoření nevelkého dialogu mezi robotem a počítačem. K tomu účelu bylo využito dvou programů, které byly poskytnuty školitelem: systém pro syntézu řeči „Epos“ a hlasový rozpoznávací software „DUNDIS“. Tento jednoduchý dialogový systém zajistil reprodukovatelnost dialogů pro různé typy bezdrátových zařízení, která mají mikrofón a reproduktor. V závěru práce byly provedeny rozsáhlé testy dialogů, ze kterých je patrný vliv použití standardních bezdrátových hlasových technologií na délku trvání dialogu a tím i rychlost reakce, například, operátora na chování robota.

Klíčová slova:

- Hlasová syntéza
- Rozpoznávání řeči
- Bluetooth
- Wifi
- Radiové spojení



ANNOTATION

The aim of this work was to study the possibility of voice communication between human and robot, determine the best way to transfer audio data over wireless networks and determine their delay compared to directly communicate without a wireless transmission. Was used to explore the three types of wireless technologies - radio connection, Wifi and Bluetooth. The work was aimed at measuring the delay, how much time elapses when transmitting audio data over wireless networks using them primarily in robotic applications. Primarily, it was necessary to create an application that recognizes human speech and if it finds an adequate response to the voice command in the vocabulary of words, further steps voice dialogue. This led to the creation of a small dialogue between the robot and the computer. For this purpose we used two programs that have been provided to the supervisor: a system for speech synthesis "Epos" and voice recognition software "DUNDIS". This simple dialogue system ensure reproducibility dialogues for different types of wireless devices that have a microphone and speaker. In conclusion, tests were conducted extensive dialogue from which it is evident the effect of using the standard wireless voice technology on the duration of the dialogue, and hence rate of reaction, for example, the operator of the robot's behavior.

Key words:

- Speech synthesis
- Speech recognition
- Bluetooth
- Wifi
- Radio link



OBSAH**STRANA**

1. ÚVOD.....	14
2. Hlasová syntéza a rozpoznávání řeči	16
2.1 Úvod do hlasové syntézy a rozpoznávání řeči	16
2.2 Dialogové systémy.....	17
2.2.1 Definice dialogu.....	17
2.2.2 Dialogový systém.....	17
3. Software	19
3.1 Hlasový rozpoznávací software „DUNDIS“	19
3.1.1 Koncepce DSR klient – server	19
3.1.2 Architektura serveru	20
3.1.3 DSR klient.....	21
3.2 Systém pro syntézu řeči „Epos“	22
3.2.1 Systém pro syntézu řeči	22
3.2.2 Architektura	23
4. Bezdrátové sítě a bezdrátové zařízení	25
4.1 Bluetooth.....	25
4.1.1 Technické řešení	25
4.2 Wifi	27
4.2.1 Standard IEEE 802.11a.....	27
4.2.2 Standard IEEE 802.11b.....	27
4.2.3 Standard IEEE 802.11g.....	27
4.2.4 Standard IEEE 802.11n.....	28
4.3 Princip radiového přenosu	28
4.4 Creative HS-1200 Digital Wireless	30
4.4.1 Technické specifikace	30
4.5 SuperTooth Buddy-Bluetooth HF.....	31
4.5.1 Vlastnosti Supertooth Buddy	31



4.6 Air Live WL-350HD	31
5. Vytvoření aplikace a provedení testování.....	33
5.1 Aplikace	33
5.2 Testování.....	36
5.3 Výsledky testování.....	39
5.3.1 Porovnání získaných údajů, přímý přístup – Bluetooth	40
5.3.2 Porovnání získaných údajů, přímý přístup – radio spojení	41
6. ZÁVĚR.....	46
7. POUŽITÁ LITERATURA	48



SEZNAM ZKRATEK A SYMBOLŮ

aj.	a jiné
atd.	a tak dále
cit.	citace
CMS	Content Management System
dBm	decibel - součinitel výkonu v decibelech
dB	decibel - jednotka intenzity zvuku
DSR	Distributed Speech Recognition
EDR	Enhanced Data-Rate
GHz	gigahertz - jednotka frekvence
GPS	Global Positioning System
GUI	Graphical User Interface
Hz	hertz - jednotka frekvence
IEEE	The Institute of Electrical and Electronics Engineers
IP	Internet Protocol
kHz	kilohertz - jednotka frekvence
L2CAP	Logical Link Control and Adaptation Protocol
LAN	Local Area Network
m	metr - jednotka délky
mm	milimetr - jednotka délky
Mbit/s	megabity za sekundu – jednotka rychlosti přenosu dat
MIMO	multiple input multiple output
MFCC	Mel-Frequency Cepstral Coefficient
mW	milliwatt - jednotka výkonu
Obr.	obrázek
Pa	pascal – fyzikální jednotka tlaku
Tab.	tabulka
t.j.	to je
TCP	Transmission Control Protocol
TTS	Text-to-Speech
TTSCP	Text-to-Speech Control Protocol



USB	Universal Serial Bus
v.	verze
VB	Visual Basic
W3C	World Wide Web Consortium
Wifi	Wireless Fidelity
WPAN	Wireless Personal Area Networks



SEZNAM OBRÁZKŮ

Obr. 2.1: Schéma základních komponent dialogového systému	18
Obr. 3.1: Bloková schéma „DUNDIS“ DSR platformy	20
Obr. 3.2: Bloková schéma DSR serveru s možným paralelním procesem rozpoznání úloh	21
Obr. 3.3: TTS systém	23
Obr. 4.1: Základní koncepce Bluetooth zařízení	26
Obr. 4.2: Radiová přenosová cesta	29
Obr. 4.3: Creative HS-1200 Digital Wireless	30
Obr. 4.4: SuperTooth BUDDY- Bluetooth HF	31
Obr. 4.5: Air Live WL-350HD	32
Obr. 5.1: Bloková schémata aplikace pro rozpoznávání a syntézu řeči	33
Obr. 5.2: Vývojový diagram dialogu	34
Obr. 5.3: Interface aplikace pro rozpoznávání a syntézu řeči	35
Obr. 5.4: Slovník pro robota	36
Obr. 5.5: Slovník pro počítač	36
Obr. 5.6: Schéma testování s použitím zařízení Creative HS-1200 Digital Wireless ...	37
Obr. 5.7: Schéma testování s použitím zařízení SuperTooth BUDDY- Bluetooth HF ..	37
Obr. 5.8: Schéma testování s použitím zařízení Air Live WL-350HD	38
Obr. 5.9: Schéma testování bez veškerého zařízení (přímý přístup)	38
Obr. 5.10: Zvukový záznam dialogu ve formě signálu	39
Obr. 5.11: Signály přímého přístupu a Bluetooth	40
Obr. 5.12: Signály přímého přístupu a radiové spojení	42
Obr. 5.13: Srovnávací analýza získaných údajů	45



SEZNAM TABULEK

Tab. 4.1: Třídy Bluetooth modulů dle komunikačního dosahu	26
Tab. 4.2: Přenosové rychlosti podle standardů	26
Tab.5.1: Délka zvukového záznamu přímého přístupu a Bluetooth.....	41
Tab. 5.2: Délka zvukového záznamu přímého přístupu a radiového spojení.	43
Tab. 5.3: Odchylka od průměru pro přímý přístup	43
Tab. 5.4: Odchylka od průměru pro Bluetooth	44
Tab. 5.5: Odchylka od průměru pro radiové spojení	44



1. ÚVOD

Vzhledem k tomu, že jedním z cílů této práce je také seznámení se současným stavem hlasového ovládání v robotických aplikacích, především na pracovišti školitele, bylo mnoho materiálů převzato z bakalářských a magisterských prací.

Problém účinného zadávání příkazů a dat do počítače je stejně starý, jako výpočetní technika sama. Způsob lidské komunikace, zcela člověku přirozený, se v podstatě pro počítače nehodí. Zatímco lidé se většinou dorozumívají pomocí mluvené řeči nebo písma (příčemž obsah dialogu a jeho význam většinou závisí na kontextu a tentýž obsah je zpravidla možno vyjádřit mnoha způsoby), počítač musí pracovat s přesně určenými elektrickými signály. Z tohoto důvodu musí být vstupní informace nejprve převedeny do formy počítači srozumitelné [1].

Do prvních počítačů zadávala údaje výhradně k tomu určená vyškolená obsluha prostřednictvím přepínačů nebo děrné pásky. S dalším rozvojem počítačů se jejich nedílnou součástí stala klávesnice a obrazovka, později i myš nebo obdobné ukazovací zařízení. Ovládání počítače bylo tak pro lidi více srozumitelné. V současné době, kdy je standardním vybavením počítače operační systém s grafickým uživatelským rozhraním, zvládne jeho obsluhu i laik [1].

Tím, že úloha převodu informace do vhodné formy byla přenechána počítači, vzrostly nároky na jeho výpočetní výkon. Není přehnané tvrdit, že při běžném domácím použití počítače se většina výpočetního výkonu spotřebuje na zpracování vstupu od uživatele a vykreslení výsledků, pouze malá část slouží k řešení samotné úlohy [1].

V současnosti se ve specializovaných oblastech začíná uplatňovat ovládání počítače pomocí hlasu. Jak bylo zmíněno na začátku, tato forma komunikace je pro člověka nejpřirozenější. Uspokojivé se zdá být vytvoření přepisu mluvené řeči do podoby textu a ovládání počítače pomocí omezeného souboru hlasových příkazů. Ovládání počítače pomocí zcela přirozené mluvy vyžaduje, aby počítač chápal význam jednotlivých slov, tento úkol je dost složitý, a proto toto zatím nebylo plně realizováno.

Možnost hlasového ovládání je atraktivní mimo jiné i v oblasti mobilní robotiky. Mobilní robot je sice možno ovládat na dálku např. pomocí notebooku a bezdrátové sítě,



ale schopnost zasahovat do jeho činnosti pomocí několika hlasových povelů přesto představuje určitý přínos [1].

Pro představu takového řešení je k dispozici klientský terminál, který rozpoznává a reprodukuje lidskou řeč, který je napojen na bezdrátovou síť ke své základně. I při využití tohoto terminálu člověkem je nutné vědět, kolik času uplyne konkrétně na přenos údajů po bezdrátových sítích. Je možné, že to nebude mít velký význam ve všech oblastech, ale existují specifická místa, kde čas hraje nemalou roli. V tom případě je důležité, kolik sekund nebo milisekund vyžaduje přenos údajů. Pak je nutné vědět, kterou bezdrátovou síť je nutné použít.

Tato diplomová práce se zabývá otázkou možnosti využívání bezdrátových technologií v komunikaci s robotem. Důležitým předpokladem k řešení této práce bylo vytvoření dialogu na základě rozpoznávání a syntézy řeči.

Pro získání nevelkého dialogu byla vytvořena aplikace, která rozpoznává a reprodukuje řeč. Pro vytvoření této aplikace, byly použity dva programy, které byly poskytnuty školitelem: systém pro syntézu řeči „Epos“ a hlasový rozpoznávací software „DUNDIS“. Potom byla aplikace instalována do dvou počítačů, na jednom z nich bylo postupně zapojeno zařízení s třemi typy bezdrátových systémů: radiové spojení, Wifi, Bluetooth. Tato zařízení mají mikrofon i reproduktor. Při spuštění aplikace na dvou počítačích, zpočátku jeden reprodukuje text, potom jej druhý rozpozná a hledá příslušnou odpověď na tento text ve své slovní zásobě. Jakmile najde odpověď, ihned reprodukuje, a tak dále. Tímto způsobem byl vytvořen nevelký dialog. Test byl proveden přímo, bez jakéhokoliv doprovodného zařízení, aby se mohly porovnat výsledky.

Diplomová práce je tématicky rozdělena na čtyři části, v první části je popsána práce systémů rozpoznávání a syntéza řeči v aplikacích pro roboty. Také je popsán systém práce programů „Epos“ a „DUNDIS“, které byly použity k vytvoření aplikace.

V druhé části najdeme popis bezdrátových systémů, jejich technické charakteristiky. Dále také popis zařízení, které byla využita k provedení testů. V třetí části je popsána činnost a práce, která byla nutná provést k vytvoření aplikace rozpoznávání a syntézy řeči. V závěrečné části je provedena analýza získaných výsledků.



2. Hlasová syntéza a rozpoznávání řeči

2.1 Úvod do hlasové syntézy a rozpoznávání řeči

Již na počátku vývoje počítače musel člověk vymyslet, jakým způsobem se s ním bude komunikovat. V počátečních fázích vývoje se převážně používaly dřené štítky, později se pro zadávání dat začala používat klávesnice a výsledek se zobrazoval na monitoru. S postupem času ke klávesnici ještě přibyla počítačová myš a grafické uživatelské rozhraní (GUI – Graphical User Interface). Dnes se vyvíjí a již postupně začíná používat v praxi další způsob komunikace člověka s počítačem – ovládání počítače pomocí hlasu. Technologie, které umožňují lidem komunikovat s počítači pomocí hlasu, mohou být rozděleny do tří hlavních kategorií:

- Hlasová syntéza (Speech synthesis) je umělá reprodukce lidského hlasu. Tato technologie je potřebná k tomu, aby počítač mohl uživateli odpovědět lidským hlasem. Systémy TTS (Text-to-Speech) převádějí běžný jazyk z jeho zaznamenané textové podoby do hlasového projevu. V současné době jsou tyto technologie využívány v praxi například na nádražích, kde příjezdy a odjezdy vlaků místo výpravčího hlásí počítač.

- Rozpoznávání řeči (Speech recognition) je technologie, která převádí lidskou řeč do textové podoby. Tato technologie je naprosto nutná pro možnost použití hlasového vstupu k ovládání počítače. V dnešní době mají technologie určené k rozpoznávání řeči pro praktické využití značné rezervy. Prozatím se v praxi používají systémy s dost omezenou slovní zásobou (hlasové ovládání GPS navigace, hlasové vytáčení, hlasové ovládání některých počítačových aplikací atd.) nebo systémy, které úspěšně dokáží rozpoznávat řeč za určitých podmínek (například tichá kancelář nebo laboratorní prostory) pouze u konkrétní osoby, na kterou byly kalibrovány. Mezi tyto aplikace patří zejména různé diktovací systémy, kde uživatel k psaní dokumentu místo klávesnice používá svůj hlas.

- Technologie pro tvorbu dialogových aplikací. Dialogová aplikace - na rozdíl od běžných aplikací - nemá grafické uživatelské rozhraní, ale komunikuje s uživatelem prostřednictvím dialogu – uživatel pomocí dialogu aplikaci přikáže, co si přeje provést, aplikace pomocí gramatiky rozezná a provede zadaný příkaz a uživateli sdělí výsledek.



Pro tvorbu dialogových aplikací se převážně používají standardy W3C Voice Browser [2].

2.2 Dialogové systémy

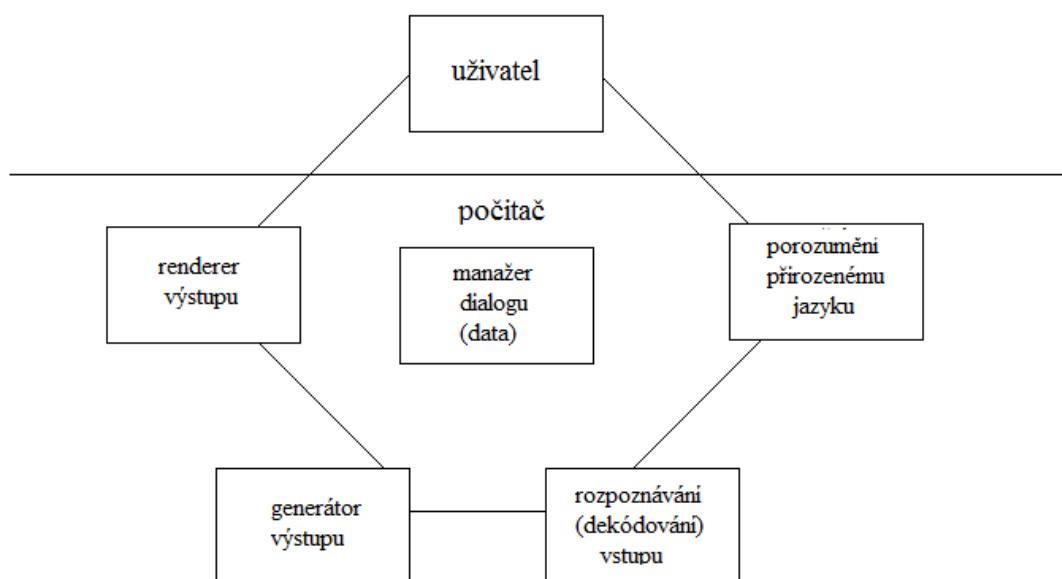
2.2.1 Definice dialogu

„Dialog je definován jako rozmluva mezi dvěma subjekty (účastníky dialogu). Jednotlivá sdělení těchto subjektů se nazývají promluvy. Tato práce uvažuje pouze dialog typu člověk - počítač. Jednotlivé promluvy na sebe plynule navazují. Pokud se dialog primárně soustředí k nalezení společného cíle, kvůli kterému je veden, je označován jako kooperativní dialog“ [4].

2.2.2 Dialogový systém

Jako dialogový systém označujeme komplexní systém disponující prostředky hlasové komunikace, který umožňuje komunikovat s počítačem přirozenou řečí prostřednictvím dialogu. Skladba těchto dialogových systémů se různí, ale v principu by měly všechny tyto systémy obsahovat tyto části:

- Porozumění přirozenému jazyku.
- Rozpoznávání/dekódování vstupu.
- Generátor výstupu.
- Renderer výstupu.



Obr. 2.1: Schéma základních komponent dialogového systému [4, s. 9].

Kromě těchto prvků obsahuje každý dialogový systém manažer (ovladač) dialogu, který provádí a zajišťuje vhodnou dialogovou strategii podle konkrétní aplikace a uchovává aktuální stav. Celý systém dále zahrnuje interní datové proměnné, které mohou být využity pro generování výstupu. Nejpodstatnější částí systému, se kterou jako jedinou přichází uživatel do kontaktu, je hlasové dialogové rozhraní. Představuje rozhraní mezi počítačem a uživatelem, podobně jako je tomu v případě např. textového či grafického uživatelského rozhraní. Toto rozhraní určuje a zodpovídá za zpracování vstupu a generování výstupu [4].



3. Software

3.1 Hlasový rozpoznávací software „DUNDIS“

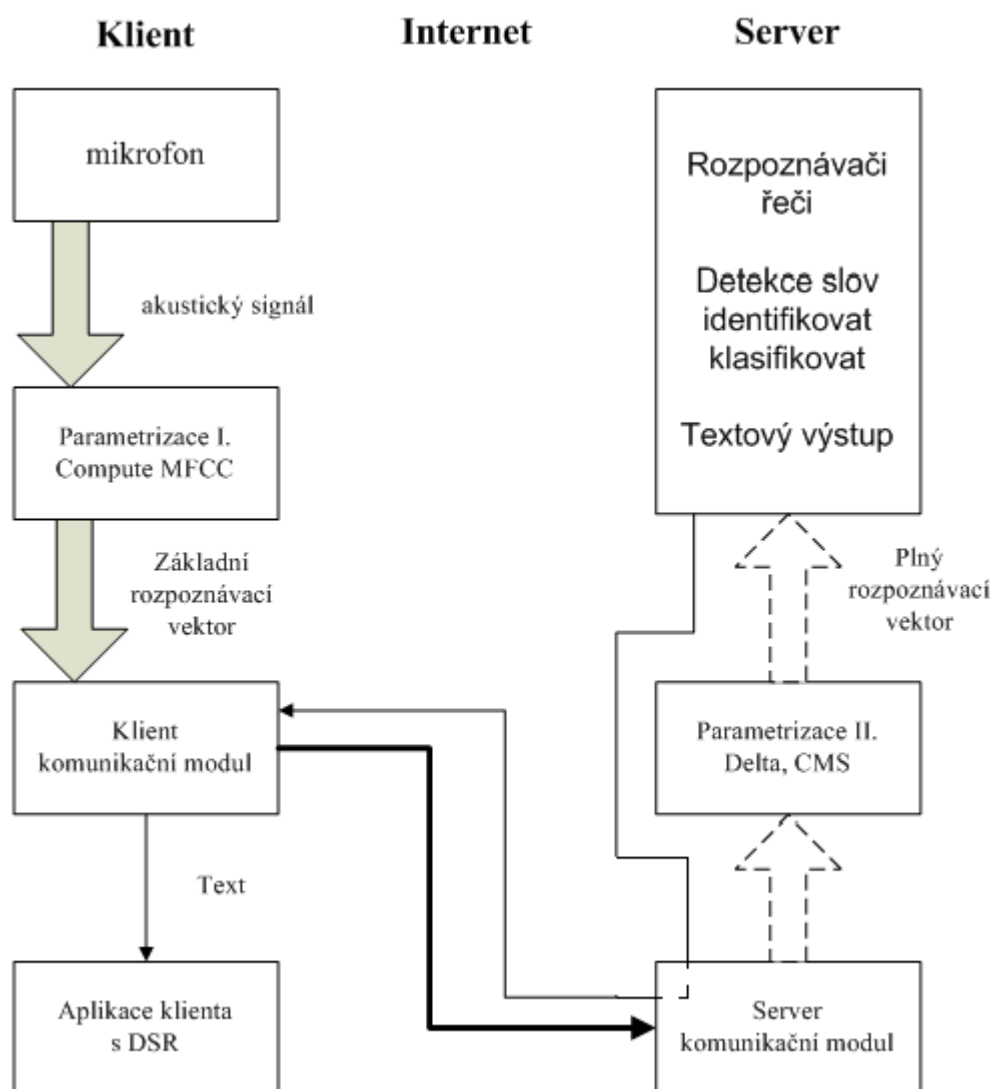
3.1.1 Koncepce DSR klient – server

Systém DSR klient – server („DUNDIS“) byl vypracován v roce 2002, tehdy hlavně pro dálkové ovládání a testování systému hlasových dialogů. Jeho schema je ukázáno na obr.3.1.

Klient–server využívá stabilní kanál přenosu dat, to garantuje komunikační protokol (například TCP/IP), že jediný správný tok údajů je získán. Tímto způsobem se nemusíme zajímat o chybné soubory.

Formát předávaných údajů je jednoduchý a lze jej lehce číst ve většině tradičních programovacích jazyků (například C, Java, VB). I když může být využíván také pro obměnu textových údajů, hlavní většina tohoto klient – serveru je sdružování i protokol předávání povelů ke startu dvojkové formy.

„DUNDIS“ se skládá z několika funkčně oddělených bloků. Ty je možné snadno spojit s modulem zpracování řeči – zajišťovat úpravy pro vytvoření konkrétního návrhu vět. Kromě toho toto modulární schématu zjednodušuje řešení a úpravy velkých systémů. Tato platforma má ještě jednu přednost: dovoluje jiným lidem (zpravidla studentům) účastnit se na výzkumu řeči bez hlubokých znalostí složitých témat, jako jsou dekodování, úprava v reálném čase a podobně.

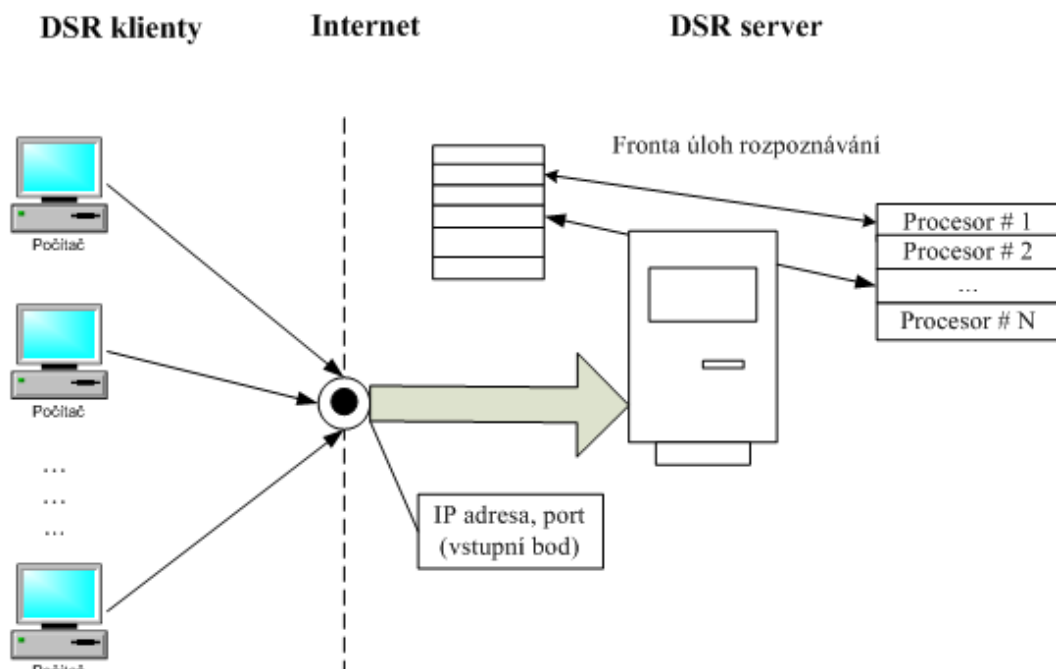


Obr. 3.1: Bloková schéma „DUNDIS“ DSR platformy [5, s. 1]

3.1.2 Architektura serveru

Schema toku údajů v serveru je ukázáno na obr.3.2. Tento systém podporuje multiprocesorové možnosti (paralelní počítač), udržovat krátkou dobu odezvy, i když jsou připojeno hodně klientů. V takovém případě řeší úkoly, které jsou předávány jednotlivými klienty, vytváří se jejich pořadí a úkoly jsou rozdělovány tak, aby se dostaly k volným procesorům. Rozpoznávání se uskutečňuje na straně serveru, podporuje se pouze diskrétní hlasový vstup (oddělená slova nebo fráze). Server může

pracovat s velkou slovní zásobou (10000 položek nebo více). Jeho akustický model byl připraven pro český jazyk.



Obr. 3.2: Blokové schéma DSR serveru s možností paralelního zpracování rozpoznání úloh [5, s. 2].

3.1.3 DSR klient

Aplikace na straně klienta je tvořena z několika modulů, které zpracovávají data na různých úrovních:

- Server – klient, nízká úroveň komunikaci – poslední verze „DUNDIS“ podporuje komunikaci založené na TCP/IP.
- „DUNDIS“- vlastní protokol – zde se využívá jednoduchý formát, který zabezpečuje předávání údajů mezi klientem a serverem. Toto dovoluje klientovi předávat binární data, jako jsou příznakové vektory nebo jejich kvantová reprezentace, a takových kontrolních dat, jako je začátek a konec rozpoznávání, určení běžné slovní zásoby atd.



- Rozhraní aplikace řeči – tento modul zahrnuje funkce dvou předchozích vrstev s tím, aby se umožnila vývojářům rychlý vývoj aplikace DSR, bez podrobné znalosti prostředků rozpoznávání řeči a při nízké úrovni komunikace.
- Uživatelské rozhraní – je jedinou aplikací, specifická část na straně klienta. To mohou být, například, okna s bohatou textovou formou při předepisování úloh [5].

3.2 Systém pro syntézu řeči „Epos“

3.2.1 Systém pro syntézu řeči

Systém pro syntézu řeči z psaného textu Epos (jinak také TTS systém z angl. Text-To-Speech) je vyvíjen jako volně šiřitelný software.

Hlavním impulsem pro jeho vývoj byla skutečnost, že v dosavadních řečových syntezech bylo nutno veškeré jazykové vlastnosti jako např. fonetickou transkripci (převod psané formy na mluvenou) a prozodická pravidla (tj. popis intonace a rychlosti řeči) psát přímo do zdrojového kódu syntezátoru. Z toho vyplývá, že prozodická pravidla mohl modifikovat v podstatě pouze autor zdrojového kódu syntezátoru.

Na základě této situace byl v r. 1996 motivován vývoj otevřeného systému pro převod psaného textu na řeč, ve kterém se jak transkripční, tak prozodická pravidla mohou zapisovat pomocí speciálního makrojazyka do konfiguračních souborů bez nutnosti znalosti zdrojového kódu a následné kompilace. Tento systém, původně určený především pro potřeby výzkumu a výuky, později dostal jméno Epos.

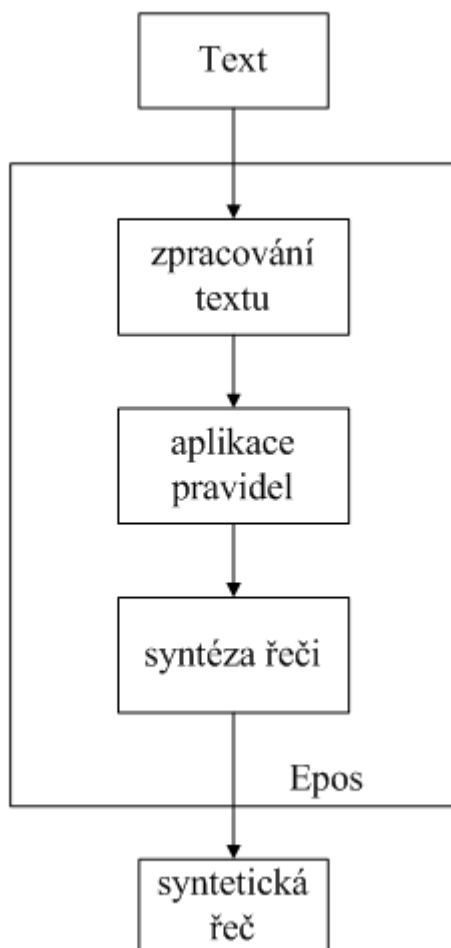
Jelikož zdrojový kód systému Epos je téměř nezávislý na syntetizovaných jazycích, je Epos vysoce konfigurovatelný. Epos umožňuje též paralelní zpracování více úloh v různých konfiguracích a v různých jazycích. K dispozici jsou nyní konfigurace pro češtinu a slovenštinu, v experimentální fázi jsou pravidla pro angličtinu, němčinu a latinu.

TTS systém Epos je založen na klient-server architektuře. Epos je možné kombinovat s několika různými synteze řeči ve frekvenční a v časové oblasti. Pro Epos byly vytvořeny jak mužské, tak ženské hlasy lišící se kvalitou i velikostí řečového inventáře. Dále je možné vybírat z několika strategií modelování prozodie:

prostřednictvím přímých pravidel, pomocí lineární predikce melodie či pomocí neuronových sítí. Po instalaci do Windows se Epos spustí jako služba TTSCP (Text-to-Speech Control Protocol).

3.2.2 Architektura

Aby se vyhnul kompromisu mezi efektivností a konfigurováním, využívá Epos model práce klient – server. Všechny potřebné procesy probíhají na serveru, takže se pojmenování Epos skutečně vztahuje k procesům serveru, a ne ke klientům.



Obr. 3.3: TTS systém.

Když bude spuštěn Epos, prochází inicializační fáze, ve které se interpretuje serveru příkazový řádek a přistupuje k souborovému systému, aby přečíst několik



konfiguračních souborů (v řádu stovek v distribuované konfiguraci), většina z tohoto procesu spočívá ve stanovení možností a analýzu transformační pravidla pro text-to-speech konverze pro každý nastavený jazyk. Epos je možné nastavit tak, aby po fázi inicializace se nevyužíval souborový systém u všech, je ve výchozím nastavení načtení několika typů informací, a ty jsou zadržovány do toho okamžiku, kdy se tato informace může využít poprvé. Téměř celá konfigurace se uskutečňuje pomocí textových souborů, které jsou transformovány do efektivních vnitřních forem.

Jakmile je ukončena inicializace, stává se Epos službou operačního systému nebo-li démonem, odposlouchávající port TCP pro nová připojení TTSCP. TTSCP představuje jediný protokol, který Epos využívá pro přenos a kontrolu informací a údajů. Neexistuje omezení ohledně množství současně připojených klientů, ani vyřizovaných dotazů, a konfigurace konstrukcí pro všechny klienty je naprosto nezávislá.

Nejčastějším úkolem Eposu je převod textu na řeč. Každý klient může akceptovat, jestli budou ve výsledku slova odesílána nazpět s využitím spojení pro předávání údajů nebo je musí server předávat přímo – bezprostředně do audio souboru operačního systému výstupu. Spolu se seřizováním a logováním je to jediná výjimka z pravidla, že TTSCP je jediným výstupním kanálem pro Epos (a v podstatě také pouze vstupní kanál) [7].



4. Bezdrátové sítě a bezdrátové zařízení

4.1 Bluetooth

Bluetooth využívá k přenosu radiové vlny vysílané v bezlicenčním pásmu 2.4 GHz, ale s menším dosahem 30 metrů ve verzi 2.1 (průmyslové moduly umožňují komunikaci až na vzdálenost několika stovek metrů). Původně byla tato technologie navržena jako náhrada kabelového spojení dvou telefonů, odtud tedy vyplývá její hlavní použití. Moduly jsou vestavěny ve většině dnešních mobilních telefonů, notebooků a dalších periferiích (headset, klávesnice, myš aj.) Specifikace definuje 29 profilů, které slouží jako podpůrné protokoly pro aplikace běžící na zařízeních, například profily pro přenos obrázků, hlasu, zpřístupnění telefonního seznamu, synchronizaci, přístup do LAN a další.

Fyzická vrstva (Radio layer) pro přenos radiových signálů využívá Frequency Hopping Spread Spectrum, které volí a využívá 79 kanálů podle naprosto náhodné frekvence, čím je na minimum minimalizována možnost rušení jinými sítěmi. Aby bylo zabráněno těmto kolizím, může slave stanice vysílat pouze, pokud je k tomu vyzvána master stanicí. Komunikace je zdvojená pomocí Time Division Duplex. Nad fyzickou vrstvou je uplatněna vrstva Baseband, která určuje dva typy logických spojů: synchronní (Synchronous Connection Oriented) pro přenos hlasu a asynchronní (Asynchronous Connection-Less) sloužící pro přenos dat. Nad touto vrstvou se nachází Link Manager Protocol, využívaný master stanicí, pečující o ustanovení spojení, autentizaci a konfiguraci spoje. Poslední vrstvou před profily a aplikační vrstvou je Logical Link Control and Adaption Protocol (L2CAP), zajišťující kvalitu služeb, multiplexing a určování – adresnost konkrétních aplikací [9].

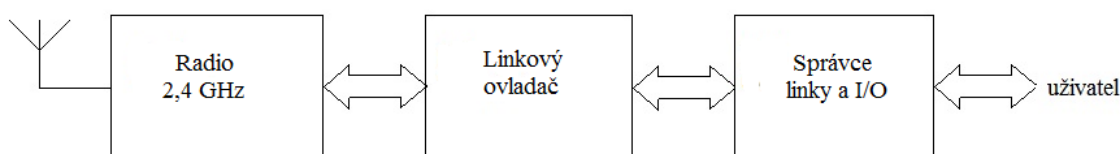
4.1.1 Technické řešení

Bluetooth systém je složen ze tří částí – komponent:

- Bluetooth radio - je ve funkci vysílač, přijímač.
- Linkový ovladač - připravuje data a zaručuje komunikaci se zařízením s Bluetooth modulem.

- Správce linky a I/O - řídí komunikaci, identifikaci, přístup a navázání spojení.

Jelikož byl Bluetooth navrhnut jako alternativa kabelu, neobsahuje žádné analogové vysokofrekvenční součástky, rozměrné a drahé komponenty jako jsou třeba filtry nebo zesilovače. Vše je ovládáno procesorem, který je schopný zpracovat i zarušený signál [10].



Obr. 4.1: Základní koncepce Bluetooth zařízení [11, s. 5].

Tab. 4.1: Třídy Bluetooth modulů dle komunikačního dosahu

Class	Maximální povolený výkon		Dosah (přibližný)
	mW	dBm	
Class 1	100	20	~100 metrů
Class 2	2.5	4	~10 metrů
Class 3	1	0	~1 metr

Zdroj: cs.wikipedia.org/wiki/Bluetooth.

Tab.4.2: Přenosové rychlosti podle standardů

Verze	Rychlost přenosu dat	Maximální propustnost
Verze 1.2	1 Mbit/s	0.7 Mbit/s
Verze 2.0 + EDR	3 Mbit/s	1.4 Mbit/s
Verze 3.0 + HS	24 Mbit/s	není stanoveno
Verze 4.0	24 Mbit/s	není stanoveno

Zdroj: cs.wikipedia.org/wiki/Bluetooth.



4.2 Wifi

4.2.1 Standard IEEE 802.11a

802.11a pracuje v licencovaném frekvenčním pásmu 5 GHz, které není tolik vytíženo a dovoluje použít více (až 8) kanálů bez vzájemného rušení. Ovšem tato pracovní frekvence nese i značné nevýhody. Ta hlavní je v nekompatibilitě s velmi rozšířenou verzí 802.11b. To byl jeden z hlavních důvodů, který zabránil masivnějšímu rozšíření 802.11a. Jinak se dá mluvit o této variantě jako o vydařeném. Oproti 802.11b byla stabilnější. Díky vyššímu vyzařovacímu výkonu a 5 GHz frekvenčnímu pracovnímu pásmu, má vyšší dosah signálu, než 802.11b. 802.11a může dosahovat teoretické rychlosti až 54 Mbit/s. Této rychlosti ale nelze běžně dosáhnout, jelikož část kapacity se spotřebuje na režii řízení. Reálná rychlost je kolem 30 Mbit/s, tedy přibližně poloviční.

4.2.2 Standard IEEE 802.11b

Jak už bylo uvedeno výše, 802.11b je nejrozšířenější variantou pro bezdrátový přenos. Mezi uživateli se rozšířila přezdívka Wifi, která následně přešla i na ostatní, později vzniklé, standardy. Pracovní frekvence je 2,4 GHz, tudíž je náchylná na rušení v již velmi zaplněném frekvenčním pásmu. Maximální rychlost na fyzické vrstvě může dosáhnout 11 Mbit/s, ale reálná rychlost je v závislosti na rušení a vlivem režie přibližně poloviční. Podle specifikace normy se rychlost na fyzické vrstvě dynamicky mění na vyšší a nižší podle toho, jak moc je prostředí pro vysílání rušivé. Rychlosti můžou nabývat hodnot 11 Mbit/s, 5,5 Mbit/s, 2 Mbit/s nebo 1 Mbit/s.

4.2.3 Standard IEEE 802.11g

Pracuje na frekvenci 2,4 GHz jako 802.11b, ale díky vylepšené technologii dosahuje rychlosti až 54 Mbit/s na fyzické vrstvě. Reálná rychlost je kolem 25 Mbit/s. Pokrok oproti 802.11b je především na fyzické vrstvě. 802.11g rozděluje data k vysílání do několika paralelních toků o nižší rychlosti. Zároveň svoje frekvenční spektrum



využívá mnohem efektivněji díky částečně se překrývajícím kanálům. Nejen díky těmto vylepšením je mnohem vyspělejší, než varianta 802.11b. Další výborná vlastnost 802.11g je její zpětná slučitelnost s 802.11b, což jí bezesporu přidává na oblibě. Ovšem v této slučitelnosti je i menší problém. Celková výkonnost 802.11g závisí především na tom, zda zrovna podporuje i klienty 802.11b. Obě varianty totiž používají odlišný způsob vysílání dat a proto nesmějí vysílat současně. Pro 802.11g to znamená další ztrátu výkonu díky vyšší režii, která řídí přizpůsobení se pomalejší 802.11b. Ale to je pouze malý nedostatek. Slučitelnost těchto dvou variant je velice vítána a nejen proto se těší velké oblibě. 802.11g patří momentálně mezi nejrychlejší bezdrátové lokální sítě pracující v otevřeném pásmu 2,4 GHz.

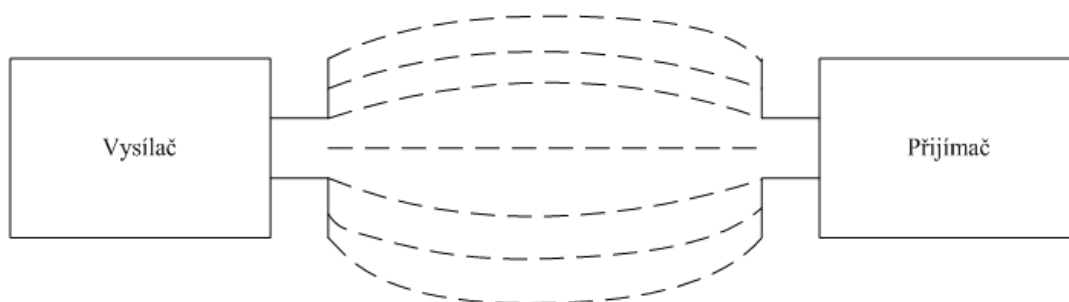
4.2.4 Standard IEEE 802.11n

Toto je zatím nejnovější norma pro bezdrátový přenos. 802.11n je velký krok vpřed v oblasti bezdrátové komunikace. Rozdíl oproti předchozím standardům je především v rychlosti. 802.11n je postavena na předchozích standardech, ale je vylepšena o řešení s názvem MIMO (multiple input multiple output). Jedná se o rozdělení původního toku dat na více toků a následné odeslání přes více antén. Následné přijetí se rovněž děje přes více antén a několik datových toků je opět spojeno v jeden. Z toho vyplývá, že čím více antén, tím vyšší rychlost. Předpokládá se až 540 Mbit/s na fyzické vrstvě. S touto rychlostí by se mohla stát velkou konkurencí kabelovým rozvodům především v domácím prostředí. 802.11n by mohla podporovat kromě počítačů, telefonů a notebooků i další multimediální přístroje, jako např. televizory, kde současné standardy kvůli nedostatečné propustnosti nemohou uspět. Samozřejmostí 802.11n bude zpětná kompatibilita s 802.11b i 802.11g. Podporovaná pracovní frekvence je 2,4 GHz i 5 GHz [12].

4.3 Princip radiového přenosu

Nízkofrekvenční signál (t.j. signál slyšitelný lidským uchem) se přemění mikrofonom na elektrické napětí. Toto napětí lze pochopitelně patřičně zesílit v napětí i výkonu a pomocí reproduktoru přivést zpět na slyšitelný zvuk.

Všichni známe zvuk šířený z reproduktorů, kdy srozumitelnost řeči je závislá na vzdálenosti a různých odrazech v prostředí reprodukce. Bezproblémově a účinně se šíří prostorem pouze vysokofrekvenční elektromagnetická vlna, t.j. vlnění se zásadně vyšším kmitočtem, než je hovorové spektrum. Proto při radiovém přenosu jsou základem vysokofrekvenční kmity, které se vhodným způsobem zesílí. Těmto kmitům říká nosný kmitočet, protože opravdu v praxi slouží jako nosič. Kmity jsou dále ovlivňovány (modulovány) slyšitelným nízkofrekvenčním signálem z mikrofону a přivedeny z vysílače do antény, která zajistí jejich šíření do okolního prostoru. Na druhé straně, někde v prostoru, je modulovaná nosná vlna zachycena přijímací anténou a v přijímači je pak vybrán signál požadovaného kmitočtu, který je dále zesílen a detekován (detekce je oddělení nízkofrekvenčního hovorového spektra od vysokofrekvenční nosné vlny). V další fázi je pak už získaný nízkofrekvenční signál zesílen a reproduktorem převedený na signál slyšitelný. Modulování nosné vlny se vytvoří pomocí modulátoru, ve kterém vzniká modulovaný signál působením nízkofrekvenčního signálu na signál nosné vlny. Používá se modulace amplitudové, frekvenční (kmitočtové) a fázové [13].



Obr. 4.2: Radiová přenosová cesta [13, s. 6]

4.4 Creative HS-1200 Digital Wireless

Bezdrátová sluchátka s mikrofonom. Pro připojení k počítači se využívá USB přijímač a komunikace mezi sluchátky a přijímačem probíhá na frekvenci 2,4 GHz [14].



Obr. 4.3: Creative HS-1200 Digital Wireless [14].

4.4.1 Technické specifikace:

Bezdrátová technologie:

- Frekvence 2,4 GHz.
- Dosah: až 22 m.

Sluchátka:

- Citlivost sluchátek: 114dB/mW
- Frekvenční rozsah: 20 Hz ~ 20 kHz.

Mikrofon:

- Citlivost mikrofону (1 kHz): -42dBV/Pa.
- Frekvence: 100 Hz ~ 16 kHz.

4.5 SuperTooth BUDDY- Bluetooth HF

Handsfree podporuje funkci Multipoint. Supertooth Buddy podporuje funkce: DSP pro potlačení okolního šumu, plně duplexní režim [15].



Obr. 4.4: SuperTooth BUDDY- Bluetooth HF [15].

4.5.1 Vlastnosti Supertooth Buddy

- Podpora DPS, full duplex.
- Bluetooth v2.1, kompatibilní s 2.0 a 1.x.
- Podporované profily: handsfree, headset.
- Operační vzdálenost až 10 metrů.
- Frekvence: 2.4 GHz.

4.6 Air Live WL-350HD

IP kamera Airlive WL-350HD. K počítači je ji možné připojit pomocí bezdrátové sítě standardu IEEE 802.11 b/g [16].

Audio:

- Plně duplexní.
- Vestavěný mikrofon.

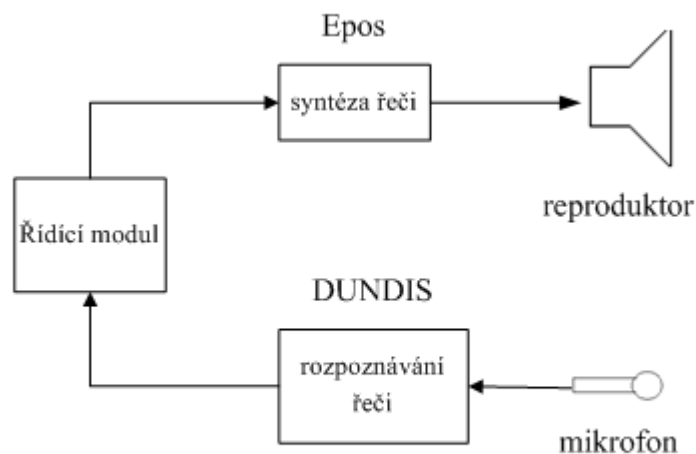


Obr. 4.5: Air Live WL-350HD [16].

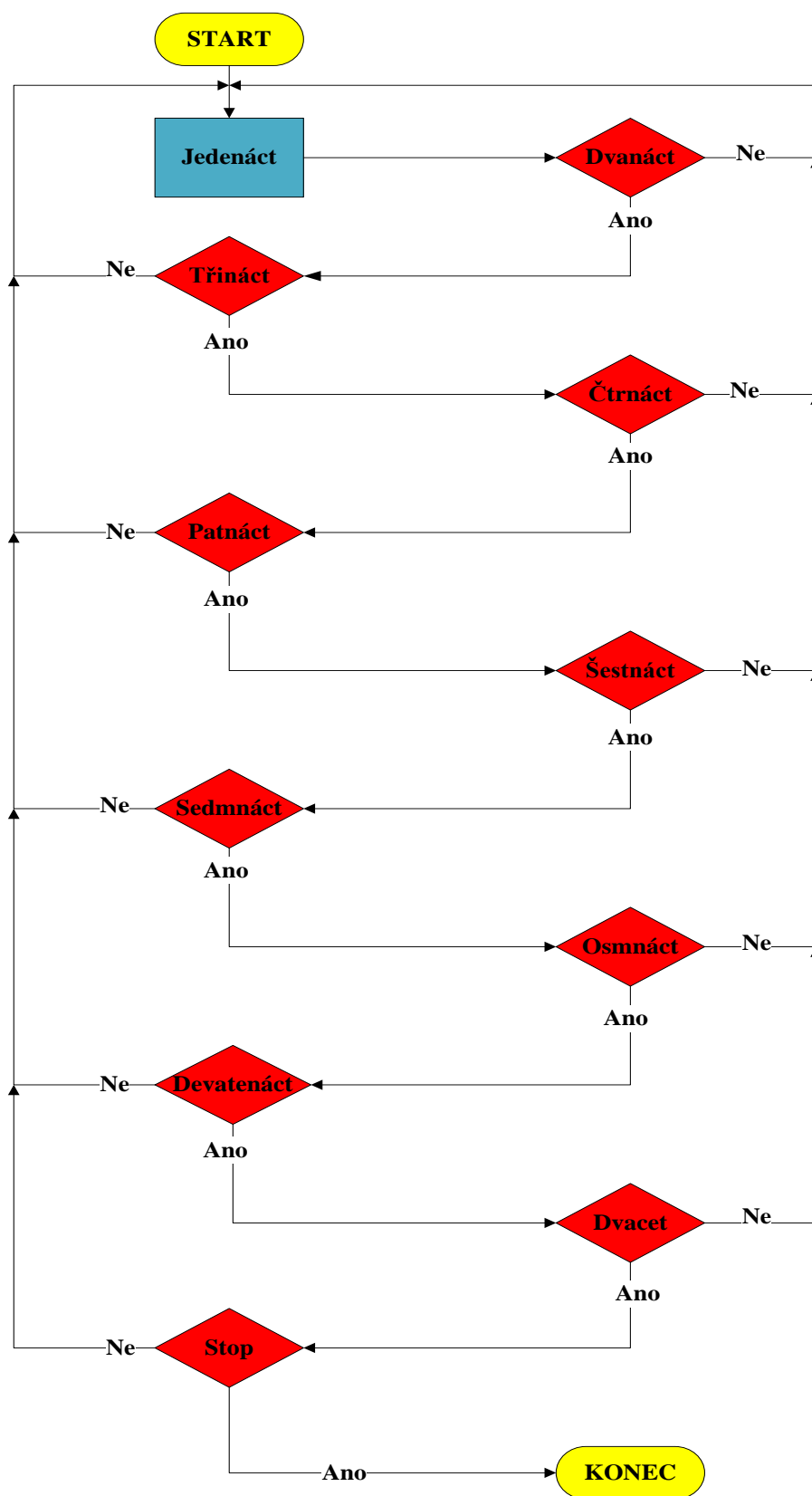
5. Vytvoření aplikace a provedení testování

5.1 Aplikace

První etapu realizace diplomové práce představovalo vytvoření aplikace pro rozpoznávání a syntézu řeči. K tomu účelu byly využity dva programy, které byly poskytnuty školitelem: systém pro syntézu řeči „Epos“ a hlasový rozpoznávací software „DUNDIS“. Systém práce aplikace byl následující: při vyslovování jakékoliv řeči do mikrofону, připojenému k počítači, rozpoznává program „DUNDIS“ tuto řeč a hledá ve speciálně zabudovaném slovníku příslušný text, odpovídající reprodukováné řeči. Pokud má tento text ve slovníku, pak vedle tohoto textu je napsán druhý text, který program Epos odesílá do klientské části, v níž existuje k tomuto textu příslušná odpověď, kterou program Epos reprodukuje (viz obrázek 5.1).



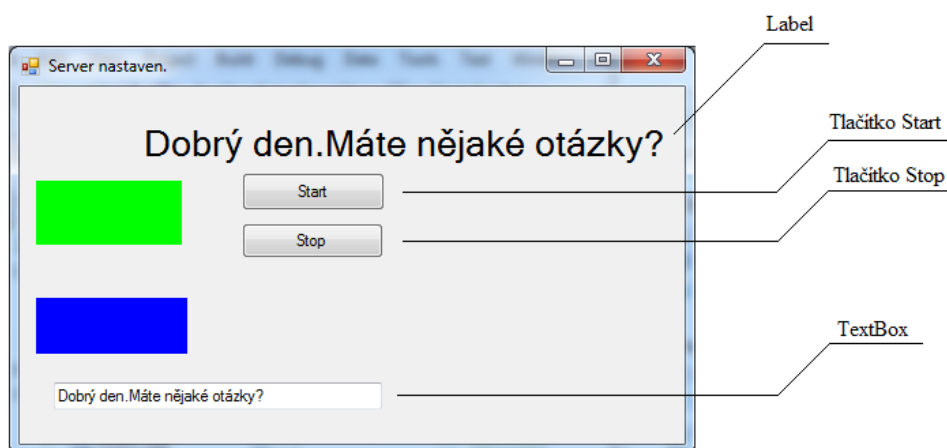
Obr. 5.1: Bloková schémata aplikace pro rozpoznávání a syntézu řeči.



Obr. 5.2: Vývojový diagram dialogu.

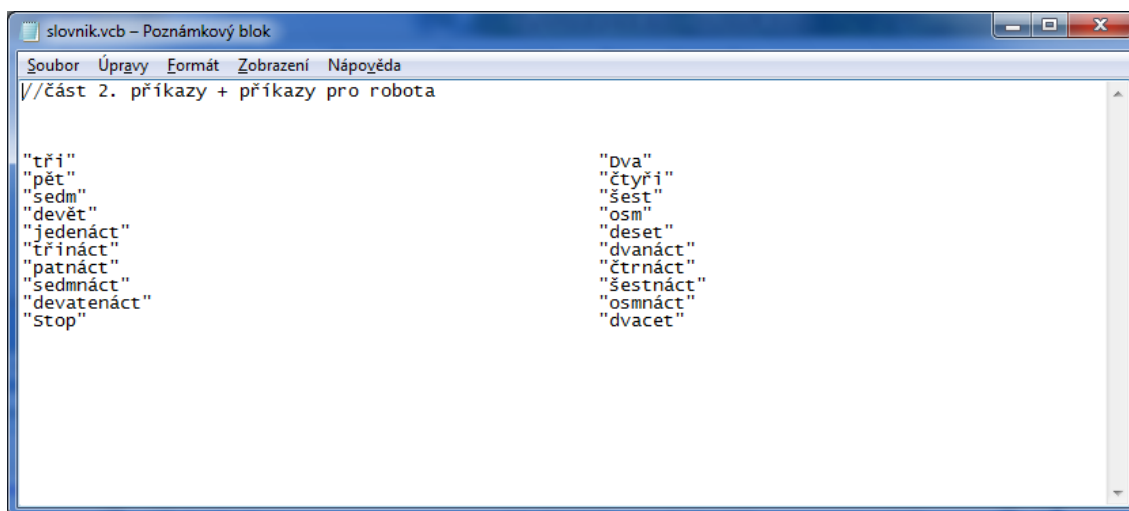
Interface této aplikace, viz obrázek 5.3, se skládá z následujících částí:

- Tlačítko Start – zapnutí aplikace.
- Tlačítko Stop – přerušení chodu aplikace.
- Label (text je nad tlačítkem Start) – je to text, který bude reprodukován.
- TextBox – k využívání pouze pro syntézu řeči z napsaného textu.

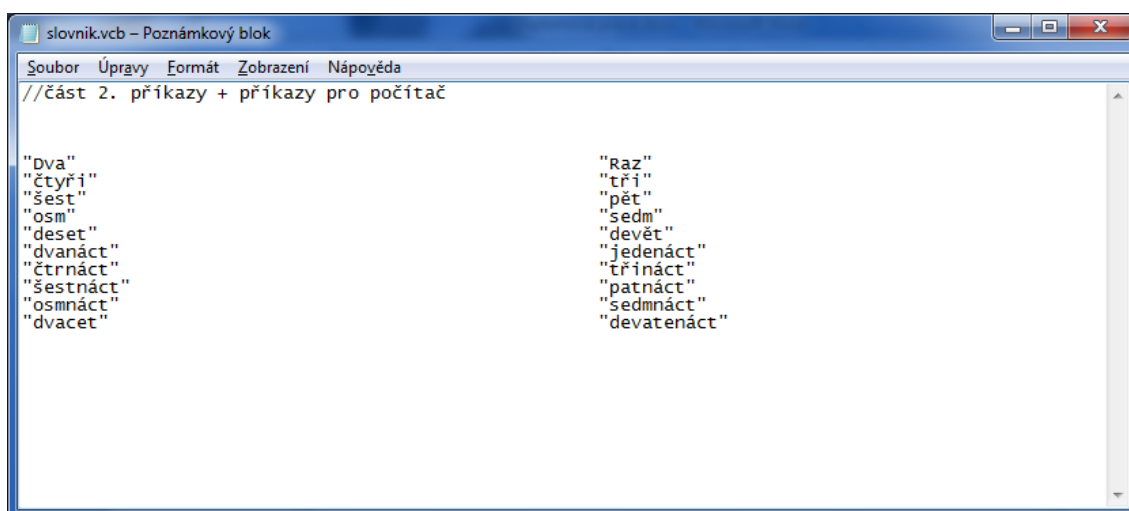


Obr. 5.3: Interface aplikace pro rozpoznávání a syntézu řeči.

Pro tvorbu dialogu byly vytvořeny dva typy slovníků, pro dva počítače. Znamená to, že text, z kterého jeden počítač provede syntézu řeči, bude rozpoznán druhým počítačem a pokud bude nalezena odpovídající odpověď v textu slovníku, z tohoto textu bude také druhým počítačem provedena syntéza řeči.



Obr. 5.4: Slovník pro robota.



Obr. 5.5: Slovník pro počítač.

5.2 Testování

Abych dostal dostatečně přesné výsledky, rozhodl jsem se provést testování, desetkrát, a to s každým typem bezdrátového zařízení.. A dále bylo také pro porovnání provedeno testování bez bezdrátového zařízení, to znamená napřímo. Při provádění těchto testů byly použity tři počítače. K jednomu počítači bylo připojeno bezdrátové zařízení, druhý počítač se stal druhou stranou dialogu, a k třetímu počítači byl připojen mikrofón a sloužil k nahrávání dialogu.

Při radiovém spojení a se zařízením Bluetooth nebyly problémy, dialog se uskutečnil. Ale při práci se zařízením Wifi, to znamená s kamerou Air Live WL-350HD vznikly komplikace. První a základní komplikací bylo zpoždění, zvukové zpoždění bylo velmi dlouhé, z toho důvodu se dialog neuskutečnil. Kdy byl spuštěn program v počítači, ke kterému byla zařazena kamera, mimo samotné kamery, také pracovali mikrofon a dynamika na počítači. To znamená, že zvuk se reprodukoval a četl v počítači i v kameře, což také rušilo průběh dialogu



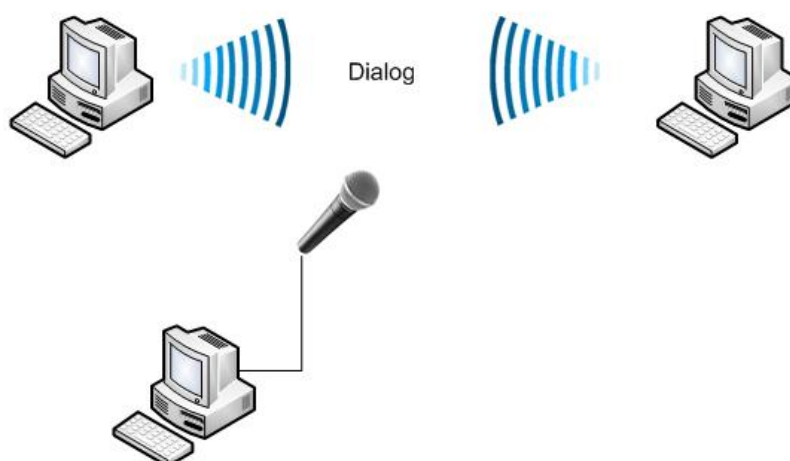
Obr. 5.6: Schéma testování s použitím zařízení Creative HS-1200 Digital Wireless.



Obr. 5.7: Schéma testování s použitím zařízení SuperTooth BUDDY- Bluetooth HF.



Obr. 5.8: Schéma testování s použitím zařízení Air Live WL-350HD.

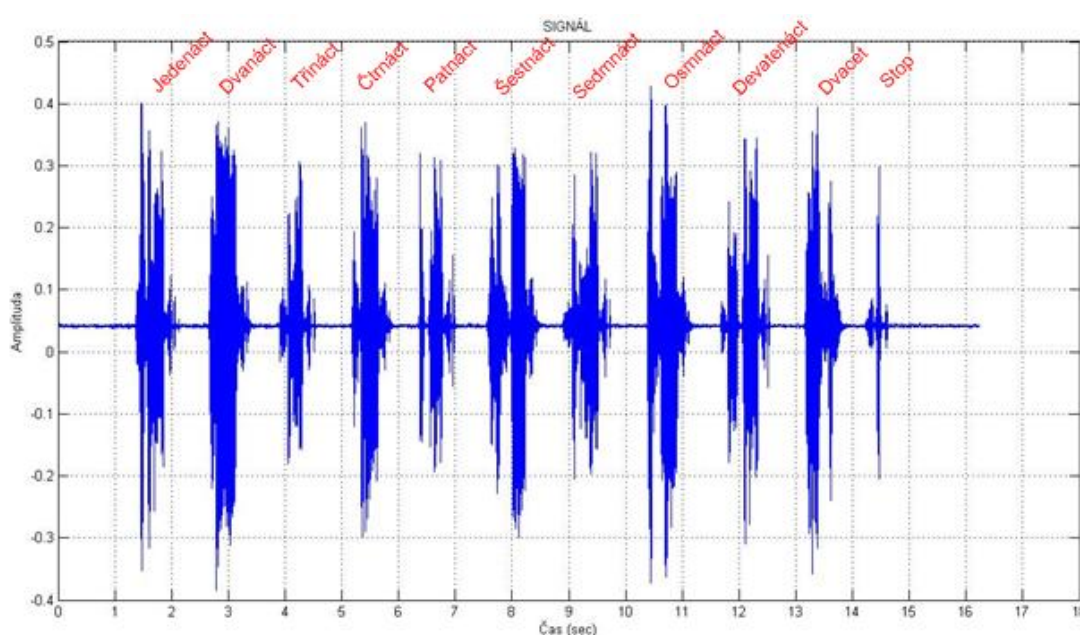


Obr. 5.9: Schéma testování bez veškerého zařízení (přímý přístup).

Každý test byl uložen ve formátu zvukového souboru wav, pro každý typ testování je provedeno 10 nahrávání. Na zvukových záznamech, při jejich reprodukování, bylo zřetelně a jasně vidět rozdíl mezi typem používaného zařízení. To znamená, že u každého typu zařízení je osobitý zvuk, což také pomáhá definovat, jaká konkrétní slova reprodukoval počítač, ke kterému bylo připojeno bezdrátové zařízení.

5.3 Výsledky testování

Aby bylo možné získat přesné údaje a určit nejlepší variantu z předkládaných návrhů bezdrátových technologií, byly všechny zvukové záznamy zpracovány v Matlabu. Tyto zvukové záznamy byly uvedeny v grafech, formou signálu, jehož vodorovná hodnota vyjadřuje čas ve vteřinách, svislá hodnota vyjadřuje amplitudu. Na obrázku 5.10 je vidět, která část signálu je to nebo ono slovo.

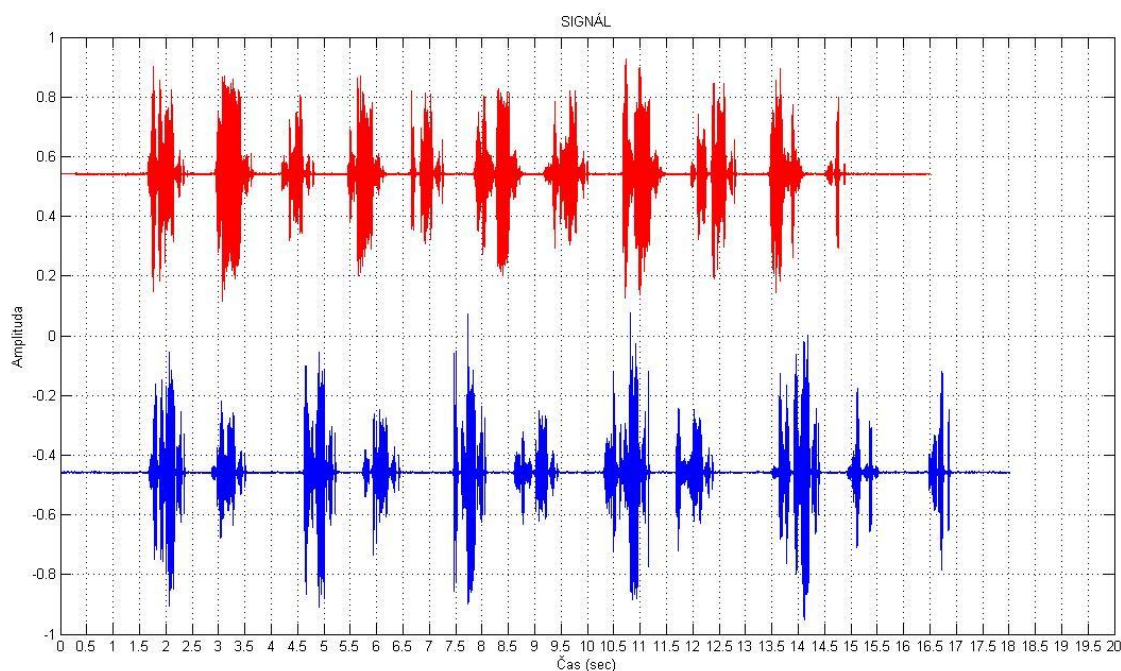


Obr. 5.10: Zvukový záznam dialogu ve formě signálu

Bylo nutné porovnat získané údaje v časovém intervalu. K tomuto účelu, získané a zpracované v Matlabu údaje, byly uvedeny v jednom grafu, a porovnávají následující údaje: přímý přístup (bez pomocného zařízení) – Bluetooth, přímý přístup – radiové spojení. Při uložení počátečních bodů signálů ve stejném čase, bylo pak na konci těchto signálů jednoznačně vidět rozdíl mezi jednotlivými druhy spojení. Různé druhy spojení byly označeny různými barvami : přímý přístup – červeně, Bluetooth – modře, radio spojení – zeleně.

5.3.1 Porovnání získaných údajů, přímý přístup – Bluetooth

Na grafu je vidět, při uložení počátečních bodů signálů ve stejném čase, že se konečný bod dialogu při přímém přístupu odlišuje o 2 sekundy od konečného bodu dialogu při použití technologie Bluetooth.



Obr. 5.11: Signály přímého přístupu a Bluetooth (přímý přístup – červeně, Bluetooth – modře).

Tabulka vytvořená k porovnání deseti zvukových záznamů při přímém přístupu a deseti zvukových záznamů, při využití technologie Bluetooth, ukazuje průměrné hodnoty údajů a také rozdíl mezi nimi. Průměrná hodnota pro přímý přístup se rovná 13,12 sec, průměrná hodnota pro Bluetooth se rovná 15,19 sec, rozdíl mezi nimi 2,07 sec.

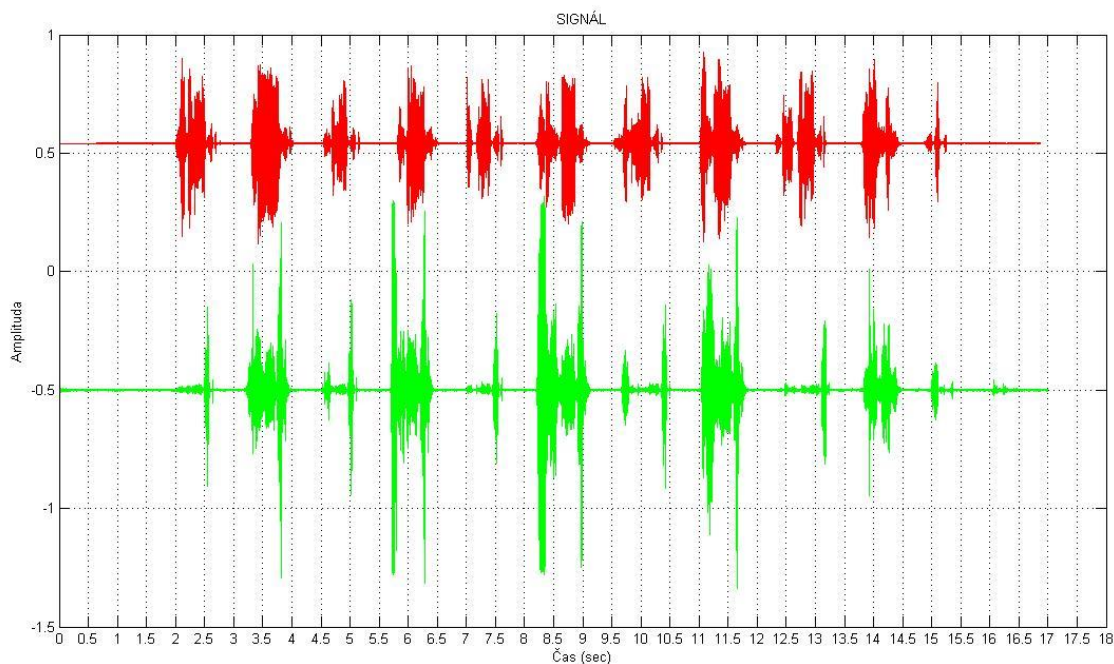


Tab.5.1: Délka zvukového záznamu přímého přístupu a Bluetooth.

№	Délka zvukového záznamu		Rozdíl (sec)
	Přímý přístup (sec)	Bluetooth (sec)	
1	13,25	15,24	
2	13,12	15,01	
3	13,21	15,62	
4	13,13	15,15	
5	13,03	15,14	
6	13,09	15,22	
7	12,97	15,34	
8	13,03	14,89	
9	13,05	15,15	
10	13,31	15,14	
Průměrná hodnota	13,12	15,19	2,07

5.3.2 Porovnání získaných údajů, přímý přístup – radio spojení

Při porovnání výsledků mezi přímým přístupem a radiovým spojením byly zjištěny rozdíly v rozsahu pouze milisekund. Lze říci, že toto spojení k přenosu dat se ukázalo jako nejlepší.



Obr. 5.12: Signály přímého přístupu a radiové spojení (přímý přístup – červeně, radio spojení – zeleně.).

Druhá tabulka vytvořená k porovnání deseti zvukových záznamů při přímém vstupu a deseti zvukových záznamů, při využití radiové spojení, ukazuje průměrné hodnoty údajů a také rozdíl mezi nimi. Průměrná hodnota pro přímý přístup se rovná 13,12 sec, Průměrná hodnota pro radiové spojení se rovná 13,20 sec, rozdíl mezi nimi 0,08 sec.

Tab. 5.2: Délka zvukového záznamu přímého přístupu a radiového spojení.

№	Délka zvukového záznamu		Rozdíl (sec)
	Přímý přístup (sec)	Radiové spojení (sec)	
1	13,25	13,30	
2	13,12	13,42	
3	13,21	13,03	
4	13,13	13,27	
5	13,03	13,18	
6	13,09	13,18	
7	12,97	12,84	
8	13,03	13,38	
9	13,05	13,42	
10	13,31	13,04	
Průměrná hodnota	13,12	13,20	0,08

Tab. 5.3: Odchylka od průměru pro přímý přístup

№	Přímý přístup (sec)	Průměrná hodnota (sec)	Odchylka od průměru (sec)
1	13,25	13,12	+0,13
2	13,12	13,12	0
3	13,21	13,12	+0,09
4	13,13	13,12	+0,01
5	13,03	13,12	-0,09
6	13,09	13,12	-0,03
7	12,97	13,12	-0,15
8	13,03	13,12	-0,09
9	13,05	13,12	-0,07
10	13,31	13,12	0,19



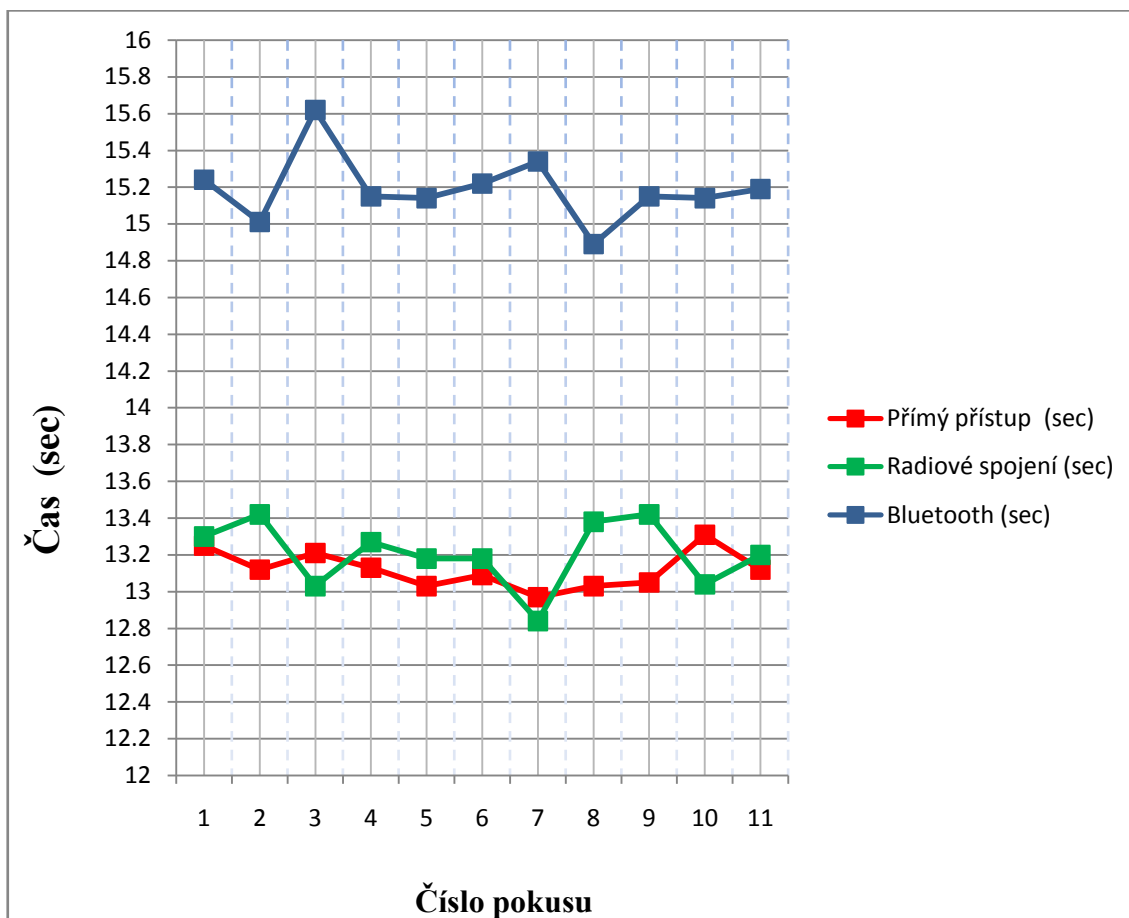
Tab. 5.4: Odchylka od průměru pro Bluetooth

№	Bluetooth (sec)	Průměrná hodnota (sec)	Odchylka od průměru (sec)
1	15,24	15,19	+0,05
2	15,01	15,19	-0,18
3	15,62	15,19	+0,43
4	15,15	15,19	-0,04
5	15,14	15,19	-0,05
6	15,22	15,19	+0,03
7	15,34	15,19	+0,15
8	14,89	15,19	-0,30
9	15,15	15,19	-0,04
10	15,14	15,19	-0,05

Tab. 5.5: Odchylka od průměru pro radiové spojení

№	Radiové spojení (sec)	Průměrná hodnota (sec)	Odchylka od průměru (sec)
1	13,30	13,20	+0,10
2	13,42	13,20	+0,22
3	13,03	13,20	-0,17
4	13,27	13,20	+0,07
5	13,18	13,20	-0,02
6	13,18	13,20	-0,02
7	12,84	13,20	-0,36
8	13,38	13,20	+0,18
9	13,42	13,20	+0,22
10	13,04	13,20	-0,16

Obr. 5.13: Srovnávací analýza získaných údajů



Z tabulek 5.3, 5.4 a 5.5, také z obrázku 5.13 je vidět, že při testování pro přímý přístup, rozptyl nebyl tak velký. Co nelze říci o výsledky získané z použití technologie Bluetooth a radiové spojení. Zde je vidět velký rozptyl, získaná data nejsou tak stabilní. Také jsou některé výsledky radiového spojení větší než výsledky přímého přístupu. 11 pokus na obrázku 5.13 – průměrná hodnota těchto údajů.



6. ZÁVĚR

Úkolem diplomové práce bylo vytvoření aplikace s možností rozpoznávání a syntézy řeči, a využít tuto aplikaci pro otestování tří typů bezdrátových zařízení: Wifi, Bluetooth a radiové spojení.

Popis možností rozpoznávání a syntézy řeči, byl uveden v druhé části této práce. Byla popsána technologie, které umožňují lidem komunikovat s počítači pomocí hlasu a také bylo uvedeno schéma základních komponent dialogového systému.

V třetí části je uveden práce systém pro syntézu řeči „Epos“ a hlasový rozpoznávací software „DUNDIS“. Jestliže je program Epos využíván pro syntézu lidské řeči, program „DUNDIS“ je využíván pro rozpoznávání řeči. Také jsou popsány charakteristiky bezdrátových systémů Wifi, Bluetooth a radiového připojení. Dále je zde uvedena obecná charakteristika zařízení, které byla využívána pro testování.

Při provádění testů byly problémy s systémem pro rozpoznávání řeči, to znamená, že aplikace ne vždy rozpoznávala slova, protože reprodukovat řeč bylo nutno provádět jasně, zřetelně. Aby systém pro rozpoznávání řeči mohl bezchybně rozpoznávat slova, byly ve vytvořeném dialogu použity delší slova, protože se krátká slova ne vždy rozpoznávají správně. A aby bylo možno vložit zvukové záznamy do grafu podle druhu signálu a následně definovat každé slovo, definovat rozdíl v časových intervalech, byly použity jedinečné slova.

Složitost problému vedla k tomu, že aplikace se komplikovaně zasekávala nebo že prostě přestávala pracovat. Zasekávání bylo způsobeno tím, že část některých slov, například koncovky, se shodovaly s jinými a to způsobilo přechod do jiné části dialogu. Abych mohl kompletně testování provést, bylo nutné zajistit důsledně i logicky syntézu a rozpoznávání těch slov, která jsou obsažena ve slovníku a která jsou nutná pro vytvoření dialogu, který by fungoval v určitou přesně danou dobu.

Ze získaných údajů lze říci, že nejlepší variantou využívání bezdrátových technologií v aplikaci pro roboty se jeví radiové spojení. Ale pokud se vezme v úvahu, že data, které se budou pomocí bezdrátových sítí přenášet v budoucnu asi porostou, a budou obsáhlejší, než bylo použito v testech, pak rozdíl v možnostech přenosu údajů bude citelný.



Také je naprosto nutné brát v úvahu tu skutečnost, že moje využívání bezdrátového zařízení bylo omezené, proto nelze pokládat tyto údaje za definitivní. je V dnešní době je dostatečně mnoho různých typů zařízení s různými možnostmi přenosu dat. Je možné provádět více podrobnějších analýz možností využívání bezdrátových technologií v aplikacích pro roboty.



7. POUŽITÁ LITERATURA

[1] ZEZULA, Miroslav. *Online detekce jednoduchých příkazů v audiosignálu* [online]. Brno: Vysoké učení technické v Brně, 2011 [cit. 2012-17-05]. Diplomová práce. Vedoucí diplomové práce Jiří Krejsa. Dostupné z:

<www.vutbr.cz/www_base/zav_prace_soubor_verejne.php?file_id=38108>.

[2] MEDŘICKÝ, Martin. *Dialogové prostředí pro vývoj dialogových aplikací* [online]. Brno: Masarykova Univerzita, 2009 [cit. 2012-17-05]. Diplomová práce. Vedoucí diplomové práce Luděk Bártek. Dostupné z: <is.muni.cz/th/99026/fi_m/diplomka.pdf>.

[3] HUANG, X., A. ACERO a H-W. HON. *Spoken language Processing, A Guide to Theory, Algorithm, and System Development*. Prentice Hall PTR, Upper Saddle River, New Jersey, 2001. ISBN 978-0130226167.

[4] BOJDA, Vojtěch. *Nástroj pro vizuální tvorbu dialogových systémů* [online]. Brno: Masarykova Univerzita, 2009 [cit. 2012-17-05]. Bakalářská práce. Vedoucí bakalářské práce Luděk Bártek. Dostupné z: <is.muni.cz/th/139813/fi_b/bp.pdf>.

[5] HOLADA, Miroslav et al. *Distributed Recognition Used as Platform for Public Testing of Speech Technology Applications* [online]. Liberec:Technická univerzita v Liberci, 2005 [cit. 2012-17-05]. Dostupné z:

<www.isca-speech.org/archive_open/archive_papers/aside_2005/aside_17.pdf>.

[6] HOLADA, Miroslav. *Design a Prototype of Client - Server Speech Recognition System*. Proc. of 6th International Workshop on Electronics, Control, Measurment and Signals-ECMS 2003. Liberec, 2003. s. 26-29. ISBN 80-7083-708-X.

[7] *The Epos Speech Synthesis System* [online]. [cit. 2012-17-05]. Dostupné z: <epos.ure.cas.cz/>.



- [8] HANIKA, Jiří a Petr HORÁK. *Text to Speech Control Protocol*. Proc. of the Int. Conf. Eurospeech'99, Budapest, Hungary, 1999, č. 5, s. 2143-2146. ISSN 1018-4074.
- [9] ČERMÁK, Milan. *Ochrana přenosu dat v přenosovém kanále systému Bluetooth* [online]. Brno: Masarykova Univerzita, 2012 [cit. 2012-17-05]. Dostupné z: http://is.muni.cz/th/325314/fi_b/bp_Milan_Cermak.txt.
- [10] STAUDEK, Jan. *WPAN, Wireless Personal Area Networks, Bluetooth* [online]. Brno: Masarykova Univerzita, 2012 [cit. 2012-17-05]. Dostupné z: www.fi.muni.cz/usr/staudek/vyuka/PA151/_wpan_bt.pdf.
- [11] KUČERA, Pavel. *Průmyslová bezdrátová síť Bluetooth pro automatizaci* [online]. Brno: Vysoké učení technické v Brně, 2006 [cit. 2012-17-05]. Dostupné z: taceo.eu/projects/golem/data/zaverecna_zprava_v8_final.pdf.
- [12] STRÁNSKÝ, Petr. *Historie Wi-Fi: od FHSS k bezdrátu* [online]. [cit. 2012-17-05]. Dostupné z: www.svethardware.cz/art_doc-E8854472EA5653EBC1257636003B03D0.html.
- [13] DVOŘÁČEK, Petr. *Radiové spojení v PO* [online]. Frýdek-Místek, 2002 [cit. 2012-17-05]. Dostupné z: www.fbi.vsb.cz/export/sites-root/fbi/030/cs/sys/resource/PDF/radiove-spojzeni.pdf.
- [14] *Creative HS-1200 Wireless Gaming Headset* [online]. [cit. 2012-17-05]. Dostupné z: us.store.creative.com/B001BAWX80/M/B001BAWX80.htm.
- [15] *SuperTooth BUDDY- Bluetooth HF* [online]. [cit. 2012-17-05]. Dostupné z: www.supertoorth.net/EN/.
- [16] *Air Live WL-350HD* [online]. [cit. 2012-30-12]. Dostupné z: cz.airlive.com/product/WL-350HD.